Empirical Study on Spatial and Temporal Features for Vehicular Wireless Communications

Yingwen Chen[1], Ming Xu[1], Pei Li[2], and Bin Zhang[3]

[1] College of Computer, National University of Defense Technology,
Changsha, Hunan 410073, China
[2] College of Information Systems and Management,
National University of Defense Technology, Changsha, Hunan 410073, China
[3] Beijing Institute of Systems Engineering, Beijing 100101, China
`csywchen@gmail.com, {xuming,peili}@nudt.edu.cn, zhangbinndu@sina.cn`

Abstract. Traditional networking technologies based on static topology analysis are not sufficient to the dynamic Vehicular Ad hoc Network. Understanding the network dynamics caused by vehicle mobility is very important for routing protocol design and algorithm optimization. This paper explores the spatial and temporal features based on two real taxi-trace datasets. It reveals that the whole topology of VANETs consists of a large number of small-sized connected components. When the communication range is greater than a threshold, a large proportion of vehicles will connect to a largest connected component, which covers the most part of the downtown region of the city both in on-peak hour and off-peak hour. Based on the analytical results, we propose several design philosophies and new research issues for VANETs.

Keywords: Wireless Communications, Spatial Temporal Analysis, Connected Component, VANETs.

1 Introduction

Vehicular Ad hoc Networking, shorten as VANET, is one kind of new technology for Intelligent Transportation Systems and Smart Cities. By using Dedicated Short Range Communications radios, vehicles can not only exchange messages directly with vicinity nodes but also communicate with other nodes through a number of intermediate nodes.

Since the nodes have capability of communicating with each other, VANETs can provide a number of potential applications with highly diverse requirements. Three major classes of applications possible in VANET are safety oriented, convenience oriented and commercial oriented[5]. Safety applications include immediate collision warning, forward obstacle detection and avoidance, emergency message dissemination and so on. Convenience applications can provide route maps with real-time traffic jams and accident conditions to help the drivers to find the shortest path in terms of time consumption. Commercial applications can provide Internet access, as well as communication between passengers in cars

Z. Cai et al. (Eds.): WASA 2014, LNCS 8491, pp. 114–125, 2014.

in the same vicinity, allowing the passengers to surfer the Internet, watch online movies, and even play games.

To support information diffusion in VANETs, two kinds of network architectures are proposed. One is called Vehicle-to-Vehicle(V2V), in which vehicles can communicate when they locate in the communication range of each other. The other one is called Vehicle-to-Roadside(V2R) or Vehicle-to-Infrastructure(V2I) or Vehicle-to-Wayside(V2W), which is a cellular-like system. The roadside infrastructure is used as a basestation and it may dominate the communication in its communication range.

In V2V VANETs, vehicles are equal and information diffusion can be achieved by adopting MANET routing protocols. However, it is well known that moving vehicles result in a disconnected network topology[14], thus different kinds of carry-and-forward strategies[14][4][8] have been proposed to support intermittent connected networks. Carry-and-forward strategy is effective but not efficient, because the source might delay the forwarding until it meets the destination even though there is a routing path between them. It is shown that the packet forwarding delay caused by carry-and-forward can be several orders-of-magnitude longer than that caused by multi-hop forwarding over a connected network[7]. Consequently, we should figure out when and where the vehicles are connected and can be reached for routing purpose.

In V2R VANETs, the roadside infrastructure is usually more powerful than vehicles. To reduce system construction cost and improve communication efficiency, the deployment of roadside infrastructures should also be well studied. Intuitively speaking, the roadside infrastructures are better setup at those spots with poor V2V connectivity.

However, the moving vehicles bring a lot of uncertainty for infrastructure deployment. No one knows whether the connectivities of vehicles are depending on the locations of the city; or whether there is any difference between on-peak hours and off-peak hours. Therefore, connectivity and time variance of network topology should be examined.

This paper is to reveal the understanding of spatial and temporal dynamics of VANETs based on two real taxi-trace datasets collected from San Francisco, USA and Shenzhen, China. The analysis results are suppose to provide new guidelines for VANETs design and protocol optimization. The main contributions of this paper are as follows:

(1) We find that by adopting a reasonable communication range, a large number of vehicles are connected as a main component of the whole network.

(2) We analyze the location dependency of the largest connected component of the whole network. We find that most part of the downtown region of the city can be covered by the largest connected component of the VANET no matter in on-peak hours or off-peak hours.

(3) According to the spatial and temporal features that we have found, we propose several design philosophies and new research issues for VANETs.

The rest of the paper is organized as follows: Section 2 briefly summarizes the related work; Section 3 describes the definitions used in spatial and temporal analysis for VANETs; Section 4 introduces the two taxi-trace datasets and presents the spatial and temporal discoveries as well as their implications for VANET design and protocol optimization. Finally, Section 5 concludes the article.

2 Related Work

With the broadcast feature of wireless channels, a VANET is always modeled by a unit disk graph, in which two vertices are connected if their distance is below a fixed threshold. By adopting this assumption, percolation theory can be used to analyze the connectivity of VANETs. For example, quantitative relationships among network connectivity, vehicle density, and transmission range are derived in [9].

When considering more realistic constraints of VANETs such as non-uniform and non-Poisson distributions, or non-ideal environments with fading/shadowing impact, Miorandi et al.[11] proposed an equivalent $GI|D|\infty$ queuing model to analyze the connectivity of one-dimensional ad hoc networks. Their results claim that no connectivity can be obtained in condition of channel randomness. Based on the equivalent $GI|D|\infty$ model, the node isolation probability and the average size of connected components can be estimated in one dimensional case. For one-dimensional VANETs, Zhuang et al.[17] also derived the exact expression for the average size of the connected components and their size distribution.

Furthermore, to analyze message propagation in two-dimensional traffic networks, the authors in [17] extended their model to calculate lattice connectivity probability for all the blocks in a district. In real traffic, most vehicles are travelling in a co-directional way. Abuelela et al. [4] found that co-directional traffic is inherently partitioned into connected components and provided an analytical expression of the expected size of those connected components. By exploring the co-directional feature and the existence of connected components, the authors in [4] also designed an Opportunistic Packet Relaying protocol (OPERA) for packet delivery over disconnected VANETs.

In recent years, complex network and network science[6] related concepts are widely adopted in many research domains. Monteiro et al.[12] decomposed the synthetic dynamical topology of VANETs into snapshots and calculated macro parameters of the network such as the node degree distribution, the clustering coefficients, the average shortest path length, and so on, for each topology snapshot.

Based on the abstracted information from network science, a new efficient broadcasting protocol called UV-CAST has been proposed. To further explore the dynamics of VANETs in a completely new way, more analysis[10][13] based on real and realistic simulated traces have been carried out in terms of many other kinds of complex network metrics. Literature [10] makes a thorough investigation of both micro-scale and macro-scale metrics including node degree, lobby index, link duration, network diameter, closeness centrality, betweenness centrality, number of communities, clustering coefficients, and so on. Similar to

[10], literature [13] conducts node-level and network-level analysis including node degree, network assortativity, betweenness centrality, and so on. Besides these, literature [13] also analyzes connected components' dynamics when the number of vehicles and communication ranges vary.

Generally speaking, literatures [9]-[17] explore VANETs' topology based on theoretical connectivity models. These models are useful in discussing the critical conditions or the connectivity boundary. But these theoretical models still rely on strong assumptions, which are very difficult to achieve in real cases. For example, the model in [9] requires a uniform distribution of the vehicles. However, it has been proved not to be true in real mobility scenarios[16]. Literatures [6]-[13] provide new methods to explore VANETs' topology based on network science models. Literatures [10] and [13] also analyze topology evolving by addressing the time-variant size of the connected components. However, their investigations have not considered the coverage area of the connected components. Different from the current work on VANETs' topology analysis, we conduct our research work based on two real taxi traces and propose a new metric to measure the location dependency of the largest connected component of the VANET.

3 Definitions

By adopting the Unit Disk Graph[3] model, VANETs topology can be abstracted as a non-directional graph tagged with timestamps. For better explanation, we first introduce some annotations for related definitions.

3.1 Network Model

The traditional static graph model in describing a network is $G =< V, E >$, where V represents the nodes and E represents the connections between the nodes. However, the VANET is dynamic and the topology is evolving due to the movement of the vehicles. Therefore, the topology of VANET can be expressed by a timestamped graph $G(T) =< V, E, T >$, where V represents all the vehicles, E represents the links between two vehicles, of which the Euclidean distance is smaller than the wireless communication range R, and T is the timestamps.

3.2 Routing Path

Routing in network is just to find a path in a given topology. We firstly define the routing path in a simplified static case. In a given timestamp t, the topology of VANET is $G(t) =< V_t, E_t, t >$. There is a non-empty sub-graph of $G(t)$, denoted as $P(t) =< V'_t, E'_t, t >$. Assume $V'_t \subseteq V, E'_t \subseteq E, n = |V'_t|, V'_t = \{v_{a1}, v_{a2}, ..., v_{an}\}$. $P(t)$ is called a *Path* if and only if there exists $\sigma : V'_t \to V'_t; \sigma(v_{ai}) = v_i (i = 1, ..., n)$,s.t. $E'_t = \bigcup_{j=1}^{n-1}(v_j, v_{(j+1)})$. Note that v_1 and v_n are called the two ends of the path at timestamp t. The path from v_1 to v_n can be denoted as Equation 1.

$$P_{v_1}^{v_n}(t) = \{v_1 \xleftrightarrow{t} v_n\} \tag{1}$$

The *length* of the path is $|V'_t|$, which is n. Note that, there are probably more than one path from v_1 to v_n in $G(t)$. Therefore, the *distance* from v_1 to v_n is defined as the shortest path from v_1 to v_n in $G(t)$. Meanwhile, there might be more than one shortest path from v_1 to v_n in $G(t)$ as well.

In a general case, routing path may pass through several timestamps. Without loss of generality, assuming the routing path from v_1 to v_n passes through a non-descending timestamp set $T_s = \left\{ t_j|_{j=1}^{j=m} \right\}$, then we have Equation 2.

$$P_{v_1}^{v_n}(T_s) = \bigcup_{j=1}^{m}(P_{v_{t_j 1}}^{v_{t_j n}}(t_j)) = \bigcup_{j=1}^{m}\{v_{t_j 1} \xleftrightarrow{t_j} v_{t_j n}\} \tag{2}$$

where $v_1 = v_{t_1 1}$,$v_n = v_{t_m n}$.

In VANET, data packets should be forwarded along the routing path in a consecutive way if the former node is connected to the latter node. However, if one node is not in the communication range of the next hop at a certain timestamp, the data packets should be buffered at this node until next timestamp comes when the consecutive two nodes on the path are connected to each other. Therefore, Equation 2 must fulfill the following requirement.

$$v_{t_k n} = v_{t_{(k+1)} 1} = V'_{t_k} \cap V'_{t_{(k+1)}} \qquad (k = 1, ..., m-1) \tag{3}$$

3.3 Connected Component

The *Connected Component* at timestamp t is a non-empty sub-graph of network $G(t)$, in which there exists at least one path for any two vertices. That is, the *Connected Component* that node v_i is connected with at timestamp t can be denoted as Equation 4.

$$CC(t) = \bigcup_{v_i, v_j \in V_t} \{v_i \xleftrightarrow{t} v_j\} \tag{4}$$

As is mentioned in section 1, packets are forwarded much faster in a muli-hop way than that in a carry-and-forward style if the source node and the destination node are in the same connected component. Therefore, the performance of the routing strategy can be greatly improved if there are enough stable connected components in VANETs. We will analyze both the number and size of the connected components in Section 4.2.

3.4 Location Dependency

The location of the connected component is also important for network design. If the connected component is location dependent, in another word, if the vehicles always form the connected component in a specified region, we do not need to place roadside infrastructures in this region anymore. In order to measure the location dependency, we need the locations of all the vehicles in the connected component. For a given connected component $CC(t)$, let $\Lambda_{CC(t)}$ represents the coordinate set of all the vehicles in $CC(t)$. Then we have $\Lambda_{CC(t)} = \{(x_{v_i}, y_{v_i}) | v_i \in V_{cc(t)}\}$. Let $\underline{X} = \min_{v_i \in V_{cc(t)}}(x_{v_i}), \underline{Y} = \min_{v_i \in V_{cc(t)}}(y_{v_i}), \overline{X} = \max_{v_i \in V_{cc(t)}}(x_{v_i}), \overline{Y} = \max_{v_i \in V_{cc(t)}}(y_{v_i})$. The rectangle covers $CC(t)$ can be defined as Equation 5.

$$\Gamma_{CC(t)} = [(\underline{X}, \underline{Y}), (\overline{X}, \overline{Y})] \tag{5}$$

where $(\underline{X}, \underline{Y})$ is the bottom left coordinate of rectangle $\Gamma_{CC(t)}$, and $(\overline{X}, \overline{Y})$ is the top right coordinate of rectangle $\Gamma_{CC(t)}$. In a consecutive timestamp set $T_s = \left\{ t_j |_{j=1}^{j=m} \right\}$, we denote the region that can cover vehicles in connected component at **any** timestamp in T_s as $\Psi = \bigcup_{j=1}^{m} \Gamma_{CC(t_j)}$, and denote the region that can cover vehicle in connected component at **all** timestamp in T_s as $\Omega = \bigcap_{j=1}^{m} \Gamma_{CC(t_j)}$. Assume the function $\delta()$ is used to calculate the area of a region, the location dependency factor of $CC(T_s)$ can be calculated by Equation 6.

$$\xi_{CC(T_s)} = \frac{\delta(\Omega)}{\delta(\Psi)} = \frac{\delta(\bigcap_{j=1}^{m} \Gamma_{CC(t_j)})}{\delta(\bigcup_{j=1}^{m} \Gamma_{CC(t_j)})} \tag{6}$$

It is obvious that $0 \leq \xi_{CC(T_s)} \leq 1$. $CC(T_s)$ is more location dependent when the value of $\xi_{CC(T_s)}$ is larger. The Location dependency of the connected components will be discussed in Section 4.3.

4 Dataset Analysis and Implications

This section we will analyze the spatial and temporal features of VANETs by using the concept of connected component other than using the individual node. Based on the discoveries, we will also present our suggestions for network design and optimization.

4.1 Taxi-Trace Dataset

As is shown in Table 1, we have collected two datasets of real taxi traces. One dataset contains GPS coordinates of more than 533 taxis collected in 20 days in San Francisco Bay area, USA[15] (SF for short). The other dataset contains GPS coordinates of 13,799 taxis in 9 days in Shenzhen, China (SZ for short). Most of the coordinate-update frequencies of SF dataset vary from 30 seconds to 60 seconds, and the SZ dataset has the coordinate-update frequency of about 30 seconds. Each vehicle has an individual trace file, in which the coordinates together with the timestamps are saved.

To capture the evolving features of network topology, we use linear interpolation to generate consecutive time synchronized coordinates with coordinate-update frequency of 30 seconds. To find the difference between topologies both in on-peak hour and off-peak hour, we select two observation intervals. One is from 1:00 am to 2:00 am, and the other is from 7:00 am to 8:00 am. To explore the variance according to different communication ranges, we choose six communication ranges from 100 meters to 600 meters.

Table 1. Taxi trace datasets. Brief summary of taxi traces for analysis.

	Taxi traces in San Francisco, USA	Taxi traces in Shenzhen, China
Vehicle quantity	533	13,799
Record duration	20 days	9 days
Update frequency	30s to 60s	30s
Taxi features	Timestamp, latitude, longitude, occupy status	Timestamp, latitude, longitude, occupy status, velocity, angle
File size	91MB compressed file	1.06GB compressed file

4.2 Size Distribution of Connected Components

Firstly, we take a look at the size distribution of the connected components, which is illustrated by Figure 1 to Figure 4. The x-axis is the fractions of the vehicles in the connected components, which is the size of the connected components divided by the total number of vehicles. The y-axis is the cumulated distribution of connected components. From the figures we can learn that most connected components are with very small size. However, the largest connected component contains a large number of vehicles. Larger communication range can enlarge the size of the connected components and hence slightly increase the fraction of big connected components. Comparing Figure 1 with Figure 2 and Figure 3 with Figure 4, we can conclude that the size distribution of the connected components is independent of the peak hours.

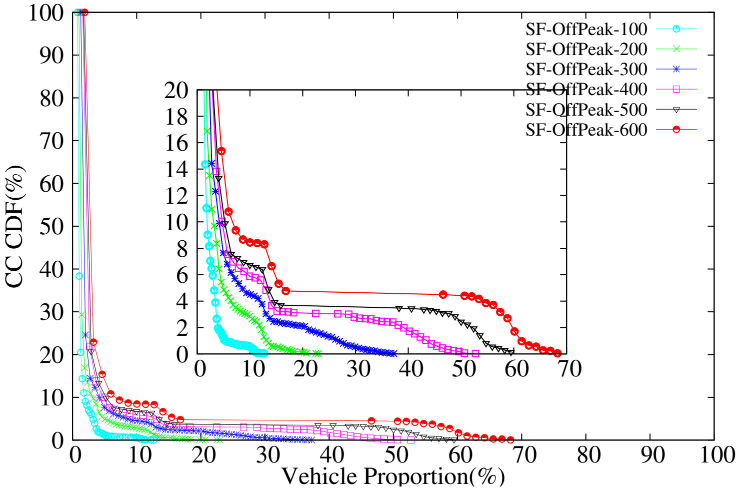

Fig. 1. Size distribution of CCs in SF during off-peak hour

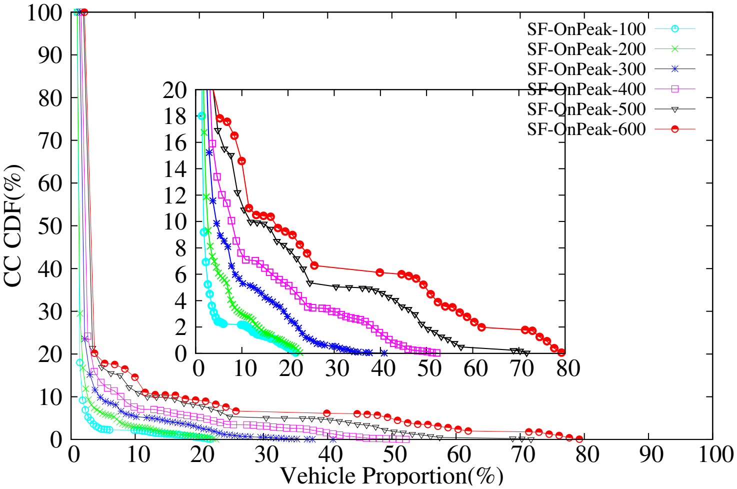

Fig. 2. Size distribution of CCs in SF during on-peak hour

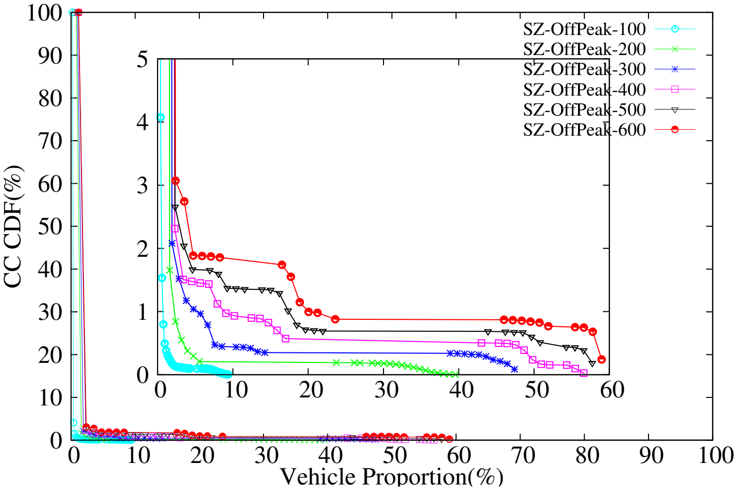

Fig. 3. Size distribution of CCs in SZ during off-peak hour

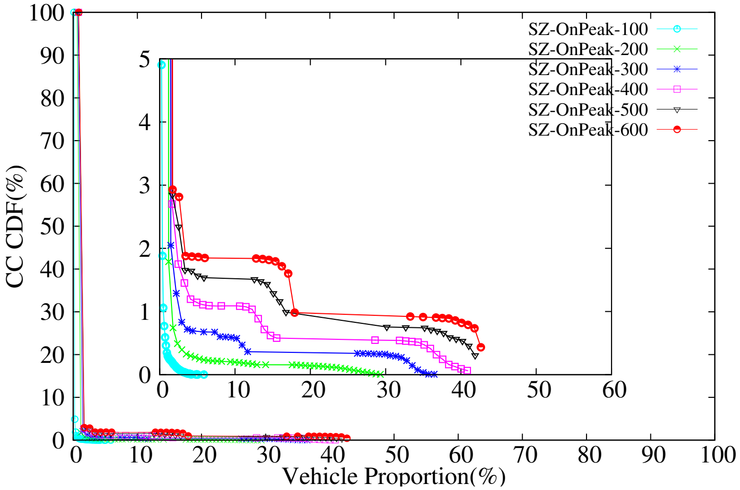

Fig. 4. Size distribution of CCs in SZ during on-peak hour.

Remark 1. Although the whole topology of the VANET is broken into a large number of small-size connected components, the largest connected component can cover a large number of vehicles. As Figure 2 shows, during the on-peak hour, the largest connected component in San Francisco even contains up to 70%-80% vehicles. Therefore, if the largest connected component keeps stable, we can make use of this feature by keeping important information on the vehicles in the largest connected component and design strategies to forward the information with the largest connected component.

4.3 Location Dependency of Connected Components

We study the location dependency of the largest connected component by calculating the location dependency factor defined in section 3.4. The results are given in Figure 5 and Figure 6. It is shown that the communication range is a critical parameter for the location dependency factor. The largest connected component covers a specific region when the communication range is large enough. In the two datasets we use, the communication range should be larger than 400 meters if we need the location dependency feature of the largest connected component. In more specified cases, when the communication range is 600 meters, we can get the results with $\delta(\Psi) \approx 56.24km^2$, $\delta(\Omega) \approx 12.21km^2$ in SF dataset and $\delta(\Psi) \approx 525km^2$, $\delta(\Omega) \approx 252km^2$ in SZ dataset. It is reported that the areas of downtown region both in San Francisco and in Shenzhen are $12.25km^2$[1] and $412km^2$[2] respectively. That means when the communication range is large enough, the largest connected component is just located in the downtown region.

As is defined in Section 3.4, $\delta(\Omega)$ is a non-increasing function according to the size of connected component, and $\delta(\Psi)$ is a non-decreasing function according to the size of connected component. Therefore, with the same communication range, location dependency factor in on-peak hour is smaller than that in off-peak hour since the largest connected component contains more vehicles in on-peak hour. However, there is a contradiction in Figure 5. That is because for unknown reasons, some of the vehicles did not successfully record their coordinates at a regular frequency (every 30-60 seconds). These vehicles are deleted from the

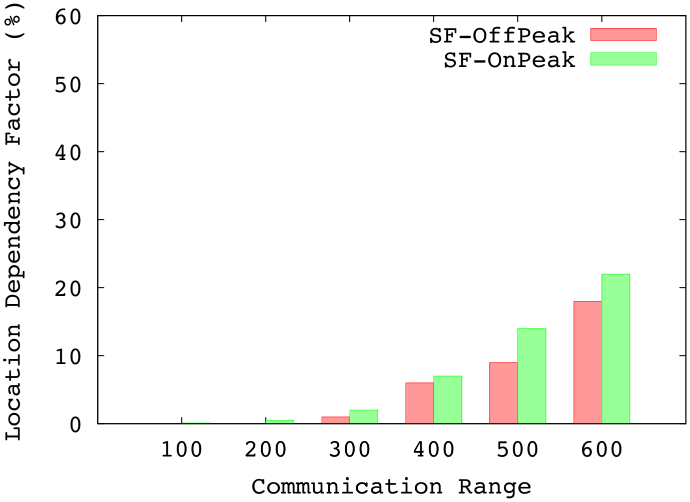

Fig. 5. Location dependency factor of largest CCs in SF

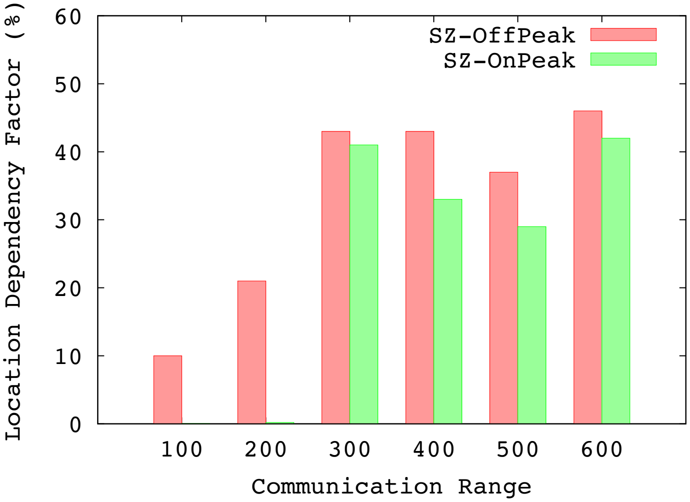

Fig. 6. Location dependency factor of largest CCs in SZ

topology due to the interpolation errors. Thus, in San Francisco case, the largest connected component contains less vehicles in on-peak hour than in off-peak hour.

Remark 2. The location dependency of the largest connected component contributes to two aspects in network design. One benefit is that we do not need to deploy roadside infrastructure to the spots where the connected component forms, because the connectivity can be maintained by the connected component. The other benefit is that vehicles in the downtown region should use multi-hop forwarding strategy rather than carry-and-forward strategy, since vehicles in this region have large probability to be connected to the largest connected component.

5 Conclusions

In this paper, we analyzed the spatial and temporal dynamics of VANETs based on two real taxi-trace datasets. We found that the whole topology of VANETs consists of a large number of small-sized connected components, however, the largest connected component among them contains a large proportion of vehicles. The performance of the routing protocol might be improved by using the largest connected component. Furthermore, by adopting a reasonable communication range, the largest connected component has the feature of location dependency, which is very useful to roadside infrastructure deployment and multi-hop packet forwarding. However, how to efficiently tracking the largest connected component in a distributed way; how to design new network architecture and routing protocols are still open research problems and they are left as our future work.

Acknowledgments. This paper is partially funded by National Natural Science Foundation of China under grant No.61003304, 61105124, and China Scholarship Council under grant 2011611534, 201206115013.

References

1. Manet driven by sf cab trace, `http://zijie.net/manet/cabtrace.php`
2. Shenzhen introduction, `http://en.wikipedia.org/wiki/Shenzhen`
3. Unit disk graph, `http://en.wikipedia.org/wiki/Unitdiskgraph`
4. Abuelela, M., Olariu, S., Stojmenovic, I.: Opera: Opportunistic packet relaying in disconnected vehicular ad hoc networks. In: 5th IEEE International Conference on Mobile Ad Hoc and Sensor Systems, MASS 2008, pp. 285–294. IEEE (2008)
5. Bai, F., Elbatt, T., Hollan, G., Krishnan, H., Sadekar, V.: Towards characterizing and classifying communication-based automotive applications from a wireless networking perspective. In: Proceedings of IEEE Workshop on Automotive Networking and Applications (AutoNet), pp. 1–25 (2006)
6. Barabasi, A.L.: Network science. Philosophical Transactions of The Royal Society 371(1987), 1–3 (2013)

7. Jeong, J., Guo, S., Gu, Y., He, T., Du, D.H.: Tsf: Trajectory-based statistical forwarding for infrastructure-to-vehicle data delivery in vehicular networks. In: 2010 IEEE 30th International Conference on Distributed Computing Systems (ICDCS), pp. 557–566. IEEE (2010)
8. Jianbin, J., Yingwen, C., Ming, X., Gu, Y.: Improving the estimation of residual delay based forwarding method in opportunistic vehicular networks. In: 2012 9th International Conference on Ubiquitous Intelligence Computing and 9th International Conference on Autonomic Trusted Computing (UIC/ATC), pp. 79–86. IEEE (2012)
9. Jin, X., Su, W., Wei, Y.: A study of the vanet connectivity by percolation theory. In: 2011 IEEE Consumer Communications and Networking Conference (CCNC), pp. 85–89. IEEE (2011)
10. Loulloudes, N., Pallis, G., Dikaiakos, M.D.: The dynamics of vehicular networks in urban environments. arXiv preprint arXiv:1007.4106 (2010)
11. Miorandi, D., Altman, E.: Connectivity in one-dimensional ad hoc networks: a queueing theoretical approach. Wireless Networks 12(5), 573–587 (2006)
12. Monteiro, R., Sargento, S., Viriyasitavat, W., Tonguz, O.K.: Improving vanet protocols via network science. In: 2012 IEEE Vehicular Networking Conference (VNC), pp. 17–24. IEEE (2012)
13. Naboulsi, D., Fiore, M.: On the instantaneous topology of a large-scale urban vehicular network: The cologne case. In: Proceedings of the Fourteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing, pp. 167–176. ACM (2013)
14. Naumov, V., Gross, T.R.: Connectivity-aware routing (car) in vehicular ad-hoc networks. In: INFOCOM 2007, 26th IEEE International Conference on Computer Communications, pp. 1919–1927. IEEE (2007)
15. Piorkowski, M., Sarafijanovic-Djukic, N., Grossglauser, M.: Crawdad trace set epfl/mobility/cab. Downloaded (February 2009), `http://crawdad.cs.dartmouth.edu/epfl/mobility/cab` (v. February 24, 2009)
16. Piórkowski, M., Sarafijanovic-Djukic, N., Grossglauser, M.: A parsimonious model of mobile partitioned networks with clustering. In: First International Communication Systems and Networks and Workshops, COMSNETS 2009, pp. 1–10. IEEE (2009)
17. Zhuang, Y., Pan, J., Cai, L.: A probabilistic model for message propagation in two-dimensional vehicular ad-hoc networks. In: Proceedings of the Seventh ACM International Workshop on VehiculAr InterNETworking, pp. 31–40. ACM (2010)